

تاریخ دریافت مقاله: ۹۸/۱۱/۲۶

تاریخ پذیرش مقاله: ۹۹/۰۴/۰۲

مدل سازی محاسباتی بازشناسی اشیاء در قشر بینایی مغز با استفاده از شبکه‌های عصبی ضربه‌ای

محمد گنج‌تابش

دانشکده ریاضی، آمار و علوم کامپیوتر - پردیس علوم - دانشگاه تهران - تهران - ایران
mgtabesh@ut.ac.ir

چکیده:

بازشناسایی اشیاء را نشان می‌دهد. همچنین ویژگی‌های استخراج شده در مدل پیشنهادی بازسازی و به همراه مکان شناسایی آن‌ها در تصاویر ورودی نشان داده شده‌اند.

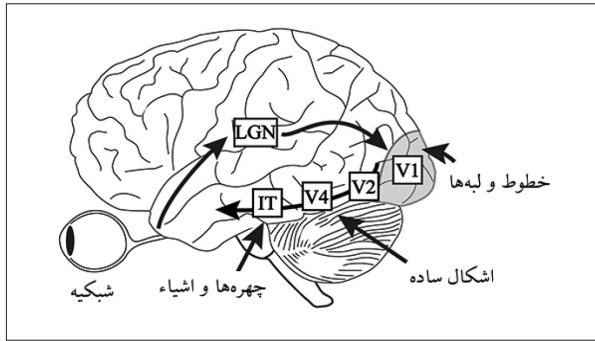
واژه‌های کلیدی: بازشناسی اشیاء، قشر بینایی، شبکه عصبی ضربه‌ای.

مقدمه

سیستم بینایی انسان قادر است تا در زمانی بسیار کوتاه و با دقتی بسیار بالا اشیاء موجود در محیط خود را در زوایا، اندازه‌ها، مکان‌ها، شرایط نوری و پس زمینه‌های مختلف شناسایی کند [۱]. با این وجود مسئله بازشناسی اشیاء از جمله پیچیده‌ترین مسائل محاسباتی در حوزه بینایی ماشین محسوب می‌گردد.

سیستم بینایی انسان از بخش‌های متعددی تشکیل شده است که فرآیند دریافت و پردازش سیگنال‌های بینایی را انجام می‌دهند (شکل ۱). در اولین مراحل از این سامانه، سیگنال‌های بینایی توسط چشم‌ها دریافت و سپس به سمت قشر بینایی در مغز به منظور ایجاد بازنمایی بهینه از تصاویر و اشیاء ارسال می‌گردد. فرآیند اصلی دریافت

تصویر منعکس شده از اشیاء پیرامون بر روی شبکه‌ی چشم همواره با تغییراتی در اندازه، زاویه دید، شرایط نوری، پس‌زمینه و نوفه همراه است. با این وجود، سیستم بینایی انسان می‌تواند اشیاء را در زمان بسیار کوتاهی شناسایی کند. تا به امروز مدل‌های محاسباتی متعددی به منظور شبیه‌سازی فرایند پردازشی سلسله‌مراتبی در قشر بینایی مغز ارائه شده‌اند که با موفقیت‌های محدودی همراه بوده‌اند. در این مقاله نشان می‌دهیم که ترکیب ساختاری مبتنی بر قشر بینایی با قواعد یادگیری موجود در مغز منجر به افزایش دقت در بازشناسی اشیاء مستقل از تغییرات می‌گردد. مدل پیشنهادی یک شبکه عصبی ضربه‌ای پیشرو است که در آن نورون‌های لایه اول لبه‌های موجود در تصویر با زوایای مختلف را شناسایی می‌کنند به طوری که نورون متناظر با لبه‌ای با تضاد نوری بیشتر سریعتر از سایر نورون‌ها فعال می‌شود. نورون‌های لایه بعدی نیز با استفاده از قاعده یادگیری وابسته به ترتیب زمانی ضربه‌ها به یادگیری ویژگی‌های بینایی موجود در تصاویر ورودی می‌پردازند. نتایج به دست آمده بر روی مجموعه داده‌های 3D-Object و ETH-80، قدرت بالای مدل پیشنهادی در



شکل ۱: مسیر بازشناسی اشیاء در قشر بینایی مغز.

گرفت، آغاز شد [۷]. مدل توصیفی پیشنهادی آن‌ها از سلول‌های ساده در ناحیه اولیه قشر بینایی مغز (ناحیه V1)، که دارای میدان دریافت‌کنندگی کوچکی بوده و به لبه‌های (خط‌های) دارای زوایای مختلف حساس هستند، شروع می‌شود و به سلول‌های پیچیده و فرای پیچیده منتهی می‌گردد. با حرکت در طول مسیر شکمی سامانه بینایی، سلول‌ها با افزایش پیچیدگی محرک‌ها و افزایش اندازه میدان دریافت همراه هستند. در انتهای مسیر شکمی در سامانه بینایی (بخش IT)، سلول‌ها برای پاسخ دادن به محرک‌های بسیار پیچیده همانند چهره تنظیم شده‌اند [۶]. ویژگی و صفت بارز سلول‌ها در ناحیه IT، توانایی بالای آنها در عملکرد مستقل از تغییرات مختلفی است که بر روی محرک‌های بینایی تاثیر می‌گذارند.

اطلاعات ارایه شده توسط دانشمندان علوم شناختی و عصبی نشان می‌دهد که فرآیند بازشناسی اشیاء در مسیر شکمی سامانه بینایی بر اصول اساسی زیر استوار است [۸]:

- ساختار سلسله مراتبی قشر بینایی در لایه‌های اولیه نسبت به موقعیت و مقیاس مقاوم بوده و در لایه‌های بعدی به زاویه دید (نقطه دید) و سایر تبدیلات پیچیده‌تر از خود مقاومت نشان می‌دهد.
- اندازه میدان دریافت‌کنندگی نورون‌ها و همچنین پیچیدگی محرک مطلوب آن‌ها با پیش رفتن در طول این سلسله مراتب افزایش می‌یابد.
- مراحل اولیه در فرآیند پردازش اطلاعات در این سیستم

سیگنال‌های بینایی توسط شبکیه^۱ چشم صورت می‌گیرد. شبکیه ناحیه حساس به نور چشم است که از سلول‌های مخروطی^۲ و استوانه‌ای^۳ تشکیل شده است که به ترتیب مسئول بینایی رنگی و سیاه سفید هستند [۲]. با تحریک سلول‌های مخروطی و یا استوانه‌ای، سیگنال‌های الکتروشیمیایی تولید و از طریق لایه‌های متوالی از سلول‌ها در شبکیه به سمت قشر بینایی در مغز ارسال می‌شوند. در ادامه سیگنال‌های بینایی به هسته خمیده جانبی^۴ در تالاموس رسیده و سپس به سمت قشر بینایی مغز فرستاده می‌شوند. در هسته خمیده جانبی دو دسته اصلی از سلول‌ها با نام‌های پاروو^۵ و ماگنو^۶ حضور دارند که سلول‌های پاروو حاوی اطلاعات جزئی‌تر بوده و در فرآیند بازشناسی اشیاء به کار می‌روند؛ در حالی که سلول‌های ماگنو در پردازش و انتقال سیگنال‌های مرتبط با مکان و یا حرکت اشیاء دخیل هستند [۳].

قشر بینایی مغز از یک ساختار لایه‌ای چندگانه تشکیل شده است. این لایه‌ها به صورت سلسله مراتبی در کنار هم قرار گرفته‌اند و اعمال محاسباتی و عصبی را بر روی سیگنال‌های بینایی انجام می‌دهند. در مراحل ابتدایی قشر بینایی، نواحی V1 و V2 عمل تشخیص لبه‌ها و خطوط را انجام می‌دهند [۴]. مرحله بعدی پردازش در ناحیه V4 صورت می‌گیرد که در آن اشیاء به وسیله ویژگی‌های پیچیده‌تری بازنمایی می‌شوند [۵]. به نظر می‌رسد که آخرین گام فرآیند بینایی، در ناحیه IT رخ می‌دهد و عمل بازشناسی اشیاء در آنجا به صورت کامل صورت می‌گیرد [۶].

مطالعات فیزیولوژیکی بر روی پستانداران (همچون گربه و میمون) چندین واقعیت مهم در مورد سازوکار بازشناسی اشیاء در قشر بینایی مغز را آشکار ساخته‌اند. این دسته از مطالعات با آزمایش‌های متنوع هابل و ویزل که بر روی گربه و سپس بر روی نوعی میمون صورت

1- Retina
2- Cone
3- Rod
4- Lateral Geniculate Nucleous (LGN)
5- Parvo
6- Magno

به صورت پیش‌رو می‌باشد (بوژه برای بازشناسی سریع اشیا).

با توجه به کارایی بسیار بالای سامانه بینایی در فرآیند بازشناسی مستقل از تغییرات اشياء، آگاهی از سازوکارهای آن در حل این مسئله از جوانب مختلف مورد اهمیت است. در واقع، دانش به وجود آمده می‌تواند به توسعه مدل‌های محاسباتی از سامانه بینایی کمک بسزایی نماید. این مدل‌های محاسباتی از سامانه بینایی می‌توانند علاوه بر استفاده در حل مسائل بینایی ماشین، به پیش‌بینی رفتار سامانه بینایی در شرایط مختلف و در مواجهه با محرک‌های مختلف و در نتیجه شناخت هرچه بهتر آن کمک نمایند. در این مقاله نشان می‌دهیم که افزودن قواعد یادگیری برگرفته از فرایند یادگیری در مغز به مدل‌هایی با ساختار منطبق بر قشر بینایی به افزایش کارایی آن‌ها در بازشناسی اشياء مستقل از تغییرات منجر خواهد شد.

۲- مدل‌سازی محاسباتی قشر بینایی مغز

تا کنون مدل‌های محاسباتی متنوعی با ویژگی‌های ساختاری و قوانین یادگیری مختلفی از عملکرد سامانه بینایی ارائه شده‌اند. مدل نئوکوگنیترون [۹] اولین مدل محاسباتی از این نوع می‌باشد که برای نخستین بار در سال ۱۹۸۰ میلادی و توسط فوکوشیما ارائه شد. این مدل دارای یک ساختار سلسله‌مراتبی است که هر بخش آن از دو لایه از سلول‌های ساده و پیچیده تشکیل شده است. سلول‌های ساده عملیات استخراج ویژگی را از طریق ترکیب ویژگی‌های به دست آمده در بخش قبلی خود انجام می‌دهند. سلول‌های پیچیده نیز با انجام عملیات میانگین‌گیری بر روی پاسخ سلول‌های ساده امکان مقاومت مکانی را برای هر ویژگی فراهم می‌کنند. در آخرین بخش از این مدل ویژگی‌های سراسری به دست آمده جهت بازشناسی الگوهای بینایی به کار می‌رود.

در سال ۱۹۹۵ میلادی، ایده شبکه‌های عصبی هم‌آمیختگی^[۱۰] که عملگر هم‌آمیختگی را به عنوان عنصر

اصلی در عملکرد سلول‌های ساده در خود داشتند توسط یان لیکن مطرح شد. این شبکه‌ها که قابلیت بالایی در استخراج ویژگی‌های تشخیصی از تصویر ورودی دارند نیز از ساختار سلسله‌مراتبی سامانه بینایی پیروی می‌کنند. این شبکه‌ها از ساختارهای دولایه متوالی تشکیل شده‌اند. این ساختارهای دولایه از یک لایه هم‌آمیختگی که امکان استخراج ویژگی در مکان‌های مختلف را حاصل می‌کند و یک لایه میانگین‌گیری که باعث به وجود آمدن بازنمایی مستقل از جابجایی می‌شوند، تشکیل شده است. در لایه‌های بالاتر، ارتباطات کامل بین سلول‌های لایه‌های متوالی باعث ترکیب ویژگی‌ها می‌شوند. در نهایت در لایه انتهایی با اعمال طبقه‌بندی نسبتاً ساده، عمل بازشناسی انجام می‌گیرد.

مدل محاسباتی HMAX برای اولین بار در سال ۱۹۹۹ میلادی ارائه شد و سپس در سال ۲۰۰۷ میلادی برای بازشناسی اشياء پیچیده به کار رفت [۱۱]. این مدل نیز از ساختار سلسله‌مراتبی قشر بینایی مغز الهام گرفته است که در آن برای یادگیری سلول‌های ساده در لایه‌های بالایی از بخش‌های بریده شده از تصویر بازنمایی شده در لایه قبلی استفاده می‌کند. به این طریق سلول‌های ساده با استخراج ویژگی‌های پیچیده، که عمدتاً بخش‌هایی از اشياء و یا کل آن‌ها را در بر می‌گیرند، توانایی بازنمایی اشياء جهت بازشناسی با دقت بالا را فراهم می‌کنند. این مدل مورد توجه بسیاری از دانشمندان علوم اعصاب بینایی بینایی ماشین قرار گرفته و از آن برای مطالعه عملکرد سامانه بینایی استفاده شده است.

در سال‌های اخیر، با پیشرفت سیستم‌های محاسباتی و همچنین در دسترس بودن مجموعه تصاویر آموزشی بسیار بزرگ، شبکه‌های عصبی هم‌آمیختگی عمیق که توسعه شبکه‌های هم‌آمیختگی هستند، توانسته‌اند به دقت بازشناسی بسیار بالایی دست یابند. در سال ۲۰۱۲ اولین مدل از این نوع توسط الکس کریژوسکی و همکاران توسعه داده شد [۱۲]. این شبکه دارای یک ساختار ۸ لایه‌ای (۵ لایه

هم آمیختگی و ۳ لایه با اتصالات کامل) است که در آن از قاعده یادگیری با ناظر پس انتشار خطا استفاده شده است. همچنین برای افزایش قدرت تعمیم، از تکنیک حذف موقت نورون‌ها در لایه‌های انتهایی استفاده شده است. موفقیت چشمگیر این شبکه، توجه محققین مختلف را به شبکه‌های عصبی عمیق جلب نمود و در سال‌های بعد شبکه‌های عمیق با کارایی بالاتر و برای کاربردهای مختلف ارائه گردیده‌اند [۱۳، ۱۴].

اگرچه شبکه‌های عصبی عمیق دقت بازشناسی بالایی دارند، اما از لحاظ شیوه انجام محاسبات و قواعد یادگیری فاصله فراوانی از شبکه‌های عصبی واقعی در مغز دارند. سلول‌های عصبی مغزی ضربه‌های الکتریکی (یا پالس‌های الکتریکی) را در سیناپس‌های خود دریافت و در صورت رسیدن به سطح پتانسیل کافی یک ضربه الکتریکی را در خروجی خود ایجاد می‌کنند. تا کنون مدل‌های نورونی مختلفی برای شبیه‌سازی دینامیک تغییرات پتانسیلی نورون‌های مغزی ارائه شده‌اند [۱۵]. مدل تجمیع و آتش نشتی^۱ ساده ترین مدل ارائه شده برای سلول‌های عصبی ضربه‌ای است [۱۶] که در آن تغییرات لحظه‌ای در پتانسیل نورون‌ها بر اساس زمان دریافت ضربه‌های الکتریکی از سایر نورون‌ها و وزن سیناپس‌های ورودی به صورت زیر تعیین می‌گردد:

$$v(t+1) = v(t) + \frac{v_{eq} - v(t)}{\tau} + I(t+1),$$

که در آن $v(t)$ اختلاف پتانسیل غشای نورون در لحظه t ، $I(t)$ جریان ورودی به نورون در لحظه t ، v_{eq} پتانسیل استراحت نورون و τ ثابت زمانی غشای نورون است که میزان تاثیر ضربه‌های دریافتی در زمان‌های مختلف را تعیین می‌کند.

شبکه‌های عصبی ضربه‌ای از اتصال نورون‌های ضربه‌ای به یکدیگر ایجاد می‌شوند. در این شبکه‌ها اطلاعات به صورت ضربه‌های الکتریکی و در طی زمان منتشر می‌شوند. تفاوت در زمان بندی ضربه‌های خروجی

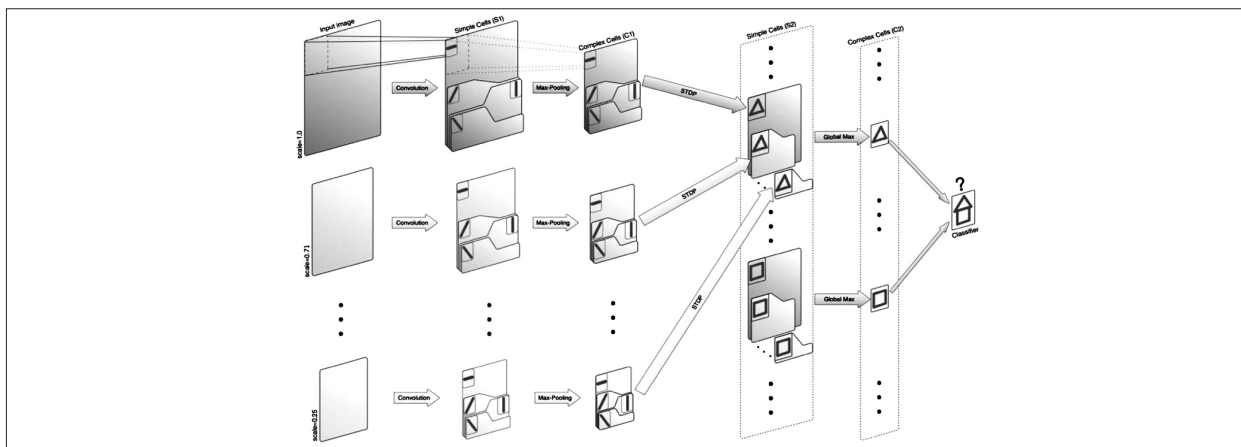
می‌تواند به عنوان کدگذاری برای اطلاعات مختلف تلقی گردد، از این رو شبکه‌های عصبی ضربه‌ای ظرفیت بالاتری در انتقال و پردازش اطلاعات نسبت به شبکه‌های عصبی عادی دارند. علاوه بر این، اتصالات نورونی دارای یک تاخیر زمانی نیز هستند که به واسطه آن ضربه‌های خروجی از نورون‌ها پس از گذشت یک زمان تاخیر به نورون‌های بعدی می‌رسند. وجود تاخیر زمانی خود باعث افزایش توانایی شبکه‌های عصبی ضربه‌ای در پردازش مسائل پیچیده می‌گردد.

در شبکه‌های عصبی ضربه‌ای از قاعده یادگیری بدون ناظر موسوم به یادگیری وابسته به زمان ضربه‌ها^۲ (STDP) استفاده می‌گردد که به زمان بندی دقیق میان فعالیت نورون‌ها بستگی دارد [۱۵، ۱۷]. قاعده یادگیری STDP بر اساس یافته‌های آزمایشگاهی و عملکرد نورون‌های مغزی به دست آمده است. در این قاعده، چنانچه ضربه ورودی از سلول پیش سیناپسی منجر به فعالیت نورون پس سیناپسی و ایجاد ضربه خروجی در آن گردد، وزن سیناپسی افزایش و در غیر این صورت کاهش می‌یابد. طبق این قاعده، اتصال سیناپسی میان نورون‌هایی که اثر بیشتری در فعال سازی یکدیگر دارند تقویت و مابقی سیناپس‌ها تضعیف می‌شوند. بدین طریق الگوهای پایدار و موثر توسط شبکه فراگرفته خواهند شد.

ماسکولیه و تورپ [۱۷] یکی از موفق ترین شبکه‌های عصبی ضربه‌ای با ساختار سلسله مراتبی الهام گرفته شده از قشر بینایی را ارائه داده‌اند. در این شبکه از مدل نورونی تجمیع و آتش استفاده شده است. همچنین این مدل از یک کدگذاری زمانی مبتنی بر یافته‌های مغزی استفاده می‌کند که در آن قسمت‌هایی از تصویر که تضاد نوری بالاتری دارند (حاوی اطلاعات بیشتری هستند) سریعتر در قشر بینایی منتشر می‌شوند. ساختار این مدل شباهت زیادی به مدل HMAX دارد، با این تفاوت که از الگوریتم یادگیری زیستی STDP برای آموزش سلول‌های ساده لایه دوم خود بهره می‌گیرد. با استفاده از این قاعده یادگیری، این

8- Leaky integrate and fire

9- Spike-timing dependent plasticity



شکل ۲: ساختار کلی مدل پیشنهادی برای بازشناسی اشیاء.

تغییرات مکانی و اندازه استفاده می‌شود. البته در این مدل از نورون‌های ضربه‌ای با عملکرد در حوزه زمان استفاده می‌شود. شمای کلی این مدل در شکل ۲ نمایش داده شده است.

زمانی که محرک ورودی به سلول‌های S1 وارد می‌شود، هرچه محرک ورودی به آن‌ها قوی‌تر و منطبق بر الگوی فراگرفته شده آن‌ها باشد، پاسخ سریع‌تری از خود در قالب یک ضربه خروجی تولید می‌کنند. سپس این ضربه‌های خروجی از S1 به صورت ناهمگام و براساس زمان ارسال آن‌ها در ساختار پیش‌رو در لایه‌های بعدی مدل انتشار می‌یابند. به عبارت دقیق‌تر، تصاویر به صورت متوالی به شبکه اعمال می‌شوند و ضربه‌های حاصل به سمت لایه S2 منتشر می‌شوند که با استفاده از قاعده یادگیری STDP به ویژگی‌های بینایی مختلف همگرا می‌شوند. سلول‌های لایه‌های S1 و S2 دارای میدان دریافت محدود بوده و از وزن‌های اشتراکی برای مجموعه سلول‌هایی که به ویژگی مشابه، اما در مکان‌ها یا اندازه‌های مختلف حساس هستند، استفاده می‌شود.

برای سلول‌های لایه S1 از فیلترهای لبه‌یاب گبور^{۱۰} در چهار زاویه مختلف استفاده شده است. این فیلترها بر روی نسخه‌هایی از اندازه‌های متفاوت (۲۵٪، ۳۰٪، ۵۰٪، ۷۱٪ و ۱۰۰٪) از تصویر ورودی اعمال می‌شوند. این امر به کاهش بار محاسباتی و همچنین افزایش مقاومت مدل

مدل به سمت یادگیری الگوهای تکرارشونده در ورودی متمایل می‌شود و بدین ترتیب ویژگی‌های تشخیصی که امکان جداسازی بهتر اشیاء از یکدیگر را دارند استخراج می‌گردند. این مدل انطباق بهتری با یافته‌های زیستی از قشر بینایی مغز داشته و عملکرد بهتری در بازشناسی اشیاء نسبت به مدل HMAX دارد.

۳- مدل پیشنهادی

در این بخش نشان می‌دهیم که چگونه افزودن قواعد یادگیری برگرفته از فرآیند یادگیری در مغز به مدل‌هایی با ساختار منطبق بر قشر بینایی می‌تواند منجر به افزایش کارایی این مدل‌ها در بازشناسی اشیاء مستقل از تغییرات شود. برای این منظور از قاعده یادگیری STDP که از فرآیند اصلاح وزن‌های سیناپسی در مغز پستانداران الهام گرفته شده است استفاده می‌گردد. سپس یک شبکه عصبی ضربه‌ای با قاعده یادگیری STDP ارائه می‌شود که قادر است تصاویر طبیعی را پردازش و نسبت به تغییرات مختلف مقاوم باشد [۱۸].

مدل پیشنهادی از یک شبکه سلسله مراتبی بهره می‌گیرد که در آن از ساختار مشابهی با مدل HMAX [۱۱] استفاده شده است. به طور مشخص، در مدل پیشنهادی نیز به صورت متوالی لایه‌ای از سلول‌های ساده، که به ویژگی‌های بینایی حساس هستند، و پس از آن لایه‌ای از سلول‌های پیچیده جهت رسیدن به مقاومت در برابر

نسبت به تغییر در اندازه اشیاء کمک می‌کند. از این رو ۲۰ دسته از سلول‌های ساده S1 وجود خواهند داشت. این سلول‌ها ضربه خروجی با میزان تاخیری متناسب با معکوس میزان تحریک ورودی از لبه متناظرشان تولید می‌کنند. توجه داشته باشید که با استفاده از قانون همه-چیز-برای-برنده^{۱۱} برای هر مکان و هر اندازه، تنها یکی از چهار سلول S1 متناظر با زوایای مختلف ضربه خروجی تولید می‌کند.

برای هر یک از دسته سلول‌های S1 که به خطوط در یک جهت و اندازه مشخص اما در مکان‌ها مختلف حساس هستند، یک صفحه سلولی C1 متناظر در نظر گرفته شده است. هر سلول C1 تنها اولین ضربه ورودی گرفته شده از یک همسایگی را به لایه بعدی منتقل می‌کند. این فرآیند معادل با عمل بیشینه‌گیری در سلول‌های پیچیده قشر بینایی مغز بوده و هدف آن دستیابی تدریجی به مقاوت به تغییرات مکانی می‌باشد.

نورون‌های لایه S2 به ویژگی‌های بینایی با پیچیدگی متوسط حساس می‌شوند که برای انجام بازشناسی اشیاء مناسب هستند. در واقع هر ویژگی در لایه S2 ترکیب وزن‌داری از خطوط استخراج شده در لایه‌های قبلی، با اندازه یکسان اما در زوایای مختلف که درون یک ناحیه همسایگی قرار گرفته‌اند، می‌باشد. هر ویژگی در لایه S2 توسط یک نورون نماینده مشخص می‌شود که مجموعه‌ای از وزن‌های سیناپسی از نورون‌های لایه C1 به آن نورون است. این نورون نماینده در مکان‌ها و اندازه‌های مختلف تکثیر می‌شود (فرآیند اشتراک وزن‌ها). این نورون‌های تکثیر شده از هر ویژگی S2، ورودی خود را از یک همسایگی در لایه C1 با اندازه متناظر با خود دریافت می‌کنند و از میان آن‌ها نورونی که زودتر به آستانه فعالیت خود برسد به عنوان نورون برنده انتخاب می‌شود. نورون برنده طی یک فرآیند مهارتی همه-چیز-برای-برنده از فعال شدن سایر نورون‌های لایه S2 با ویژگی یکسان جلوگیری می‌کند. سپس برای هر ویژگی، نورون برنده فرآیند

یادگیری STDP را برای بروز رسانی ماتریس وزن اجرا می‌کند. این تغییرات در تمام نورون‌های مشابه با نورون برنده تکثیر می‌شود. بدین صورت ویژگی‌های بینایی که بسامد بیشتری در مجموعه تصاویر ورودی داشته باشند توسط نورون‌های S2 استخراج می‌شوند.

همچنین برای جلوگیری از همگرایی نورون‌های نماینده S2 به ویژگی‌های بینایی یکسان از یک سازوکار مهار محلی بین نورون‌های نماینده استفاده می‌شود. زمانی که یک نورون S2 در مکان و اندازه مشخص فعال می‌شود، از فعالیت نورون‌های همسایه با ویژگی‌های متفاوت جلوگیری می‌کند. برای هر ویژگی S2 یک نورون C2 معادل وجود دارد که خروجی نورون‌های S2 مربوط به یک ویژگی در کلیه مکان‌ها و اندازه‌های مختلف را دریافت می‌کند. در واقع خروجی نورون‌های C2 به صورت سراسری نسبت به تغییرات مکانی و اندازه در ویژگی‌های استخراج شده مقاوم خواهند بود.

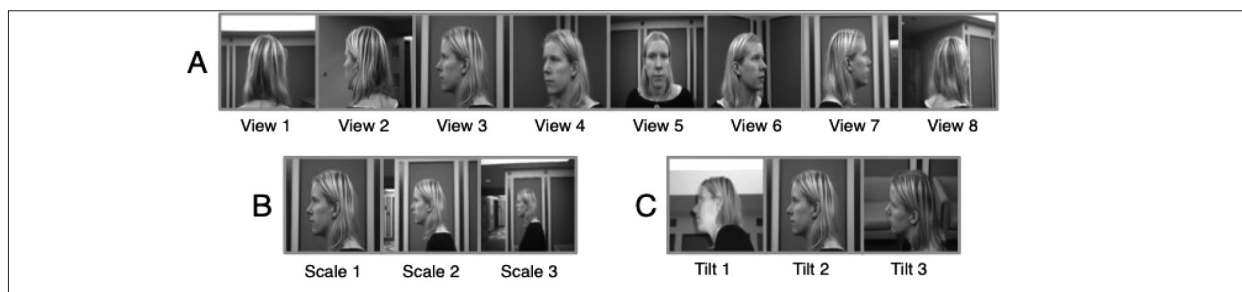
جهت یادگیری ویژگی‌های بصری در وزن‌های اتصالی بین لایه‌های C1 به S2 از قاعده ساده شده STDP به صورت زیر استفاده می‌شود:

$$w(\Delta t) = \begin{cases} A_+ e^{-\frac{\Delta t}{\tau_+}}, & \Delta t \geq 0, \\ A_- e^{-\frac{\Delta t}{\tau_-}}, & \Delta t < 0, \end{cases}$$

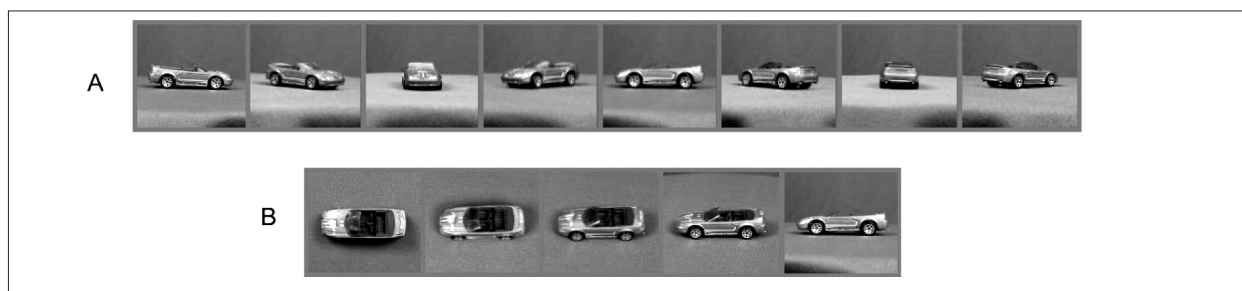
که در آن $\Delta t = t_{post} - t_{pre}$ اختلاف زمان ضربه نورون‌های پس‌سیناپسی و پیش‌سیناپسی را نشان می‌دهد، A_+ و A_- به ترتیب نرخ یادگیری برای افزایش و کاهش وزن سیناپسی هستند، و τ_+ و τ_- نشان دهنده پنجره یادگیری هستند. لازم به ذکر است که در این قاعده اختلاف زمانی بین دو ضربه تأثیری در میزان تغییر وزن نداشته و تنها علامت آن در نظر گرفته می‌شود.

همان‌گونه که قبلاً اشاره شد، فرآیند ساخت ویژگی‌ها در لایه S2 کاملاً به صورت بدون ناظر انجام می‌شود. از این رو در حین انجام یادگیری بدون ناظر STDP، تصاویر آموزشی به صورت کاملاً تصادفی و بدون در نظر گرفتن برچسب کلاس، زاویه دید و اندازه شیء به مدل اعمال

11- Winner-takes-all



شکل ۳: نمونه‌هایی از تصاویر سر انسان از مجموعه داده 3D-Object در (A) زوایا، (B) اندازه‌ها و (C) ارتفاع‌های دید مختلف.



شکل ۴: نمونه‌هایی از تصاویر ماشین از مجموعه داده ETH-80 در (A) زوایا و (B) ارتفاع‌های دید مختلف.

می‌باشد. برای هر رده ده نمونه مختلف وجود دارد که از هر کدام در ۷۲ شرایط مختلف شامل هشت زاویه دید، سه فاصله (اندازه) و سه ارتفاع دید مختلف عکسبرداری شده است. تصاویر بخش‌بندی نشده‌اند و اشیاء در پس‌زمینه‌های متفاوتی قرار گرفته‌اند. شکل ۳ تعدادی از تصاویر سر انسان که در شرایط مختلفی تصویربرداری شده است را نمایش می‌دهد. مجموعه داده ETH-80 نیز شامل هشت رده از اشیاء مختلف، شامل سیب، گلابی، گوجه‌فرنگی، ماشین، گاو، اسب، سگ و فنجان، است که هر کدام از ۴۰ نقطه دید (۸ زاویه و ۵ ارتفاع) تصویر برداری شده‌اند (شکل ۴).

کارایی مدل پیشنهادی و مدل HMAX مورد تحلیل و بررسی قرار گرفته است که نتایج آن در جدول ۱ قابل مشاهده می‌باشد. برای هر دو مدل نتایج طبقه‌بندی بر اساس تعداد ویژگی‌های مختلف محاسبه شده است. نتایج به دست آمده نشان می‌دهند که مدل پیشنهادی می‌تواند برای مجموعه داده 3D-Object تنها با استفاده از ۳۰۰ ویژگی به کارایی بالا و قابل قبول دست یابد. همچنین برای مجموعه داده ETH-80 نیز مدل پیشنهادی با تعداد ویژگی کمتر نسبت به مدل HMAX به دقت بازشناسی بالاتری

می‌شوند. فرآیند آموزش از وزن‌های تصادفی آغاز و زمانی که ۷۰۰ ضربه برای هر ویژگی ایجاد شده باشد خاتمه می‌یابد. سپس STDP غیرفعال شده و توانایی ویژگی‌ها در بازنمایی اشیاء آزمون مستقل از تغییرات بررسی می‌گردد. برای محاسبه ویژگی‌های C2 متناظر با هر تصویر آزمون، آستانه فعالیت آن‌ها را برابر با بی‌نهایت در نظر گرفته و پتانسیل نهایی آن‌ها را محاسبه کرده‌ایم. در واقع پتانسیل نهایی هر سلول برابر است با تعداد ضربه‌های متناظر با آن ویژگی در لایه S2. در نهایت یک طبقه‌بند خطی ماشین بردار پشتیبان^{۱۲} (SVM) جهت انجام فرآیند طبقه‌بندی مبتنی بر ویژگی‌های C2 استفاده شده است.

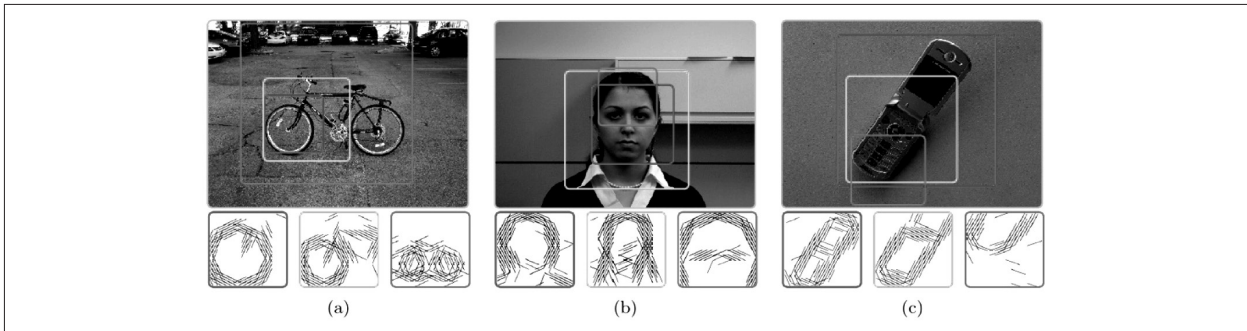
۴- نتایج

جهت بررسی قدرت مدل پیشنهادی در مواجهه با تغییراتی در اندازه، زاویه دید و پس‌زمینه، از مجموعه داده‌ی 3D-Object [۱۹] و ETH-80 [۲۰] استفاده شده است. مجموعه داده 3D-Object شامل ده رده از اشیاء مختلف، شامل دوچرخه، ماشین، تلفن همراه، سر انسان، اتو، مانیتور، ماوس کامپیوتر، کفش، ماشین منگنه و تستر نان،

12- Support vector machine

جدول ۱: دقت بازشناسی اشیاء در مدل پیشنهادی و مدل HMAX برای تعداد متفاوتی از ویژگی‌ها.

مدل HMAX				مدل پیشنهادی				مجموعه داده	
۱۲۰۰۰	۹۰۰۰	۳۰۰۰	۱۰۰۰	۵۰۰	۴۰۰	۳۰۰	۲۰۰	تعداد ویژگی‌ها	3D-Object
۶۲/۴	۶۱/۹	۶۰/۱	۵۸/۲	۹۶/۰	۹۶/۰	۹۴/۷	۷۶/۱	درصد کارایی	
۵۰۰۰	۲۰۰۰	۱۰۰۰	۵۰۰	۱۲۵۰	۱۰۰۰	۷۵۰	۵۰۰	تعداد ویژگی‌ها	ETH-80
۶۹/۰	۶۸/۹	۶۸/۷	۶۶/۳	۸۱/۱	۸۰/۷	۷۹/۳	۷۵/۳	درصد کارایی	



شکل ۵: بازنمایی سه ویژگی مختلف در لایه S2 به همراه ویژگی‌های مطلوب آن‌ها که از روی ماتریس وزن بازسازی شده‌اند.

ورودی کافی هستند. به علاوه، سلول‌های فعال شده برای تصویر ورودی کل سطح شیء هدف را پوشش داده و پاسخی به نواحی مربوط به پس‌زمینه تصاویر نمی‌دهد. در واقع، قاعده یادگیری STDP به صورت ذاتی بر روی بخش‌های مشترک تصاویر آموزشی با بسامد تکرار بالا، که همان شیء هدف می‌باشد، تمرکز می‌کند و نواحی مربوط به پس‌زمینه، که در میان تصاویر مختلف متفاوت است، به صورت کامل نادیده گرفته می‌شود.

۵- نتیجه‌گیری

تا به امروز شبکه‌های عصبی متعددی با معماری‌هایی برگرفته شده از قشر بینایی مغزی ارائه شده‌اند، اما این مدل‌ها عمدتاً فرآیند یادگیری در مغز را نادیده گرفته و از روش‌های یادگیری متفاوتی، که معمولاً توجیه زیستی برای آن‌ها وجود ندارد، استفاده می‌کنند. در این مقاله نشان دادیم که استفاده از ساختار قشر بینایی در کنار قاعده یادگیری منطبق با مغز منجر به افزایش عملکرد مدل‌ها در بازشناسی اشیاء مستقل از تغییرات می‌شود. ویژگی‌های بینایی بدست آمده از طریق قاعده یادگیری STDP، که در

دست یافته است.

مطالعات گوناگون نشان داده‌اند که مدل HMAX، به دلیل عدم دستیابی به بازنمایی مناسب و مقاوم به تغییرات، عملکرد ضعیفی در بازشناسی اشیاء در حالتی که تغییرات شدید در ورودی وجود داشته باشد دارد. شباهت ساختاری بالای مدل HMAX و مدل پیشنهادی و عملکرد بسیار ضعیف مدل HMAX در مقابل مدل پیشنهادی، گواه محکمی بر اهمیت و قدرت یادگیری بدون ناظر STDP در استخراج ویژگی‌های تشخیصی مناسب که حاوی اطلاعات زیادی هستند می‌باشد.

همانطور که گفته شد، استخراج ویژگی‌های تشخیصی در مدل پیشنهادی عامل دسترسی آن به دقت بازشناسی بالا می‌باشد. برای درک بهتر این موضوع با استفاده از یک فرآیند بازسازی سعی به بازنمایی ویژگی‌های فراگرفته شده توسط نورون‌های لایه S2 و مکان شناسایی آن‌ها در تصاویر ورودی کردیم. شکل ۵ ناحیه دریافت تعدادی از نورون‌های لایه S2 همراه با بازنمایی محرک مطلوب آن‌ها را نمایش می‌دهد. همانطور که در این شکل دیده می‌شود، تنها تعداد کمی از نورون‌های لایه S2 برای بازنمایی اشیاء

12. [12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
13. [13] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pages 818–833. 2014.
14. [14] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*, pages 487–495, 2014.
15. [15] Filip Ponulak and Andrzej Kasinski. Introduction to spiking neural networks: Information processing, learning and applications. *Acta Neurobiologiae Experimentalis*, 71(4):409–433, 2010.
16. [16] Anthony N Burkitt. A review of the integrate-and-fire neuron model: I. homo- geneous synaptic input. *Biological Cybernetics*, 95(1):1–19, 2006.
17. [17] Timothée Masquelier and Simon J Thorpe. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3(2):e31, 2007.
18. [18] Saeed Reza Kheradpisheh, Mohammad Ganjtabesh, and Timothée Masquelier. Bio-inspired unsupervised learning of visual features leads to robust invariant object recognition. *Neurocomputing*, 205:382–392, 2016.
19. [19] Silvio Savarese and Li Fei-Fei. 3d generic object categorization, localization and pose estimation. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.
20. [20] Bastian Leibe and Bernt Schiele. Analyzing appearance and contour based methods for object categorization. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–409. IEEE, 2003.

مدل پیشنهادی به کار رفته است، حاوی اطلاعات مناسب بوده و در مقابل تغییرات مختلف اشیاء بسیار مقاوم هستند. با وجودی که ساختار مدل پیشنهادی بسیار ساده بوده و تنها یک لایه از آن قابلیت یادگیری دارد و همچنین قانون یادگیری نیز در ساده‌ترین حالت خود مورد استفاده قرار گرفته است، اما نتایج به دست آمده نشان از کارایی بالای این مدل در مسئله بازشناسی اشیاء دارد. این امر حاکی از آن است که با پیچیده‌تر کردن معماری مدل و استفاده از قواعد یادگیری منطبق‌تر با مغز می‌توان کارایی مدل‌های محاسباتی را بیش از پیش بهبود بخشید.

مراجع

1. [1] Rufin VanRullen and Simon J Thorpe. Surfing a spike wave down the ventral stream. *Vision research*, 42(23):2593–2615, 2002.
2. [2] Helga Kolb, Ralph Nelson, Peter Ahnelt, and Nicolas Cuenca. Cellular organization of the vertebrate retina. *Progress in brain research*, 131:3–26, 2001.
3. [3] Neil R Carlson. *Physiology of behavior*. Allyn & Bacon, 1986.
4. [4] John HR Maunsell and William T Newsome. Visual processing in monkey extrastriate cortex. *Annual review of neuroscience*, 10(1):363–401, 1987.
5. [5] Anna W Roe, Leonardo Chelazzi, Charles E Connor, Bevil R Conway, Ichiro Fujita, Jack L Gallant, Haidong Lu, and Wim Vanduffel. Toward a unified theory of visual area v4. *Neuron*, 74(1):12–29, 2012.
6. [6] Tamas Tompa and Gyula Sáry. A review on the inferior temporal cortex of the macaque. *Brain research reviews*, 62(2):165–182, 2010.
7. [7] Margaret Livingstone and David Hubel. Segregation of form, color, movement, and depth- anatomy, physiology, and perception. *Science*, 240(4853):740–749, 1988.
8. [8] Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber, and Tomaso Poggio. Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, 29(3), 2007.
9. [9] Kunihiko Fukushima. Artificial vision by multi-layered neural networks: Neocognitron and its advances. *Neural Networks*, 37:103–119, 2013.
10. [10] Yann LeCun. Learning invariant feature hierarchies. In *European Conference on Computer Vision*, pages 496–505. Springer, 2012.
11. [11] Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber, and Tomaso Poggio. Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, 29(3), 2007.